WHAT IS CLAIMED IS:

1. A method for predicting the potential of an oligonucleotide to hybridize to a target nucleotide sequence, said method comprising:

(a) identifying a predetermined number of unique oligonucleotides within a nucleotide sequence that is hybridizable with said target nucleotide sequence, said oligonucleotides being chosen to sample the entire length of said nucleotide sequence,

SUB A3

(b) determining and evaluating for each of said oligonucleotides at least one parameter that is independently predictive of the ability of each of said oligonucleotides to hybridize to said target nucleotide sequence,

(c) identifying a subset of oligonucleotides within said predetermined number of unique oligonucleotides based on an examination of said parameter, and

(d) identifying oligonucleotides in said subset that are clustered along a region of said nucleotide sequence that is hybridizable to said target nucleotide sequence.

2. A method according to Claim 1 which comprises ranking said oligonucleotides of step (d) based on the size of said clusters of oligonucleotides.

3. A method according to Claim 1 wherein said unique oligonucleotides are of identical length N.

SUB B2

4. A method according to Claim 3 wherein said unique oligonucleotides are spaced one nucleotide apart, said predetermined number comprising L-N+1 oligonucleotides, where L is the length of the hybridizable sequence.

5. A method according to Claim 1 wherein said parameter is selected from the group consisting of composition factors, thermodynamic factors, chemosynthetic efficiencies and kinetic factors.

6. A method according to Claim 1 wherein said parameter is a composition factor selected from the group consisting of mole fraction (G+C), percent (G+C), sequence complexity, and sequence information content.

7. A method according to Claim 1 wherein said parameter is a thermodynamic factor selected from the group consisting of predicted duplex melting temperature, predicted enthalpy of duplex formation, predicted entropy of duplex formation, predicted free energy of duplex formation, predicted melting temperature of the most stable intramolecular structure of the oligonucleotide or its complement, predicted enthalpy of the most stable intramolecular structure of the oligonucleotide or its complement, predicted entropy of the most stable intramolecular structure of the oligonucleotide or its complement, predicted free energy of the most stable intramolecular structure of the oligonucleotide or its complement, predicted melting temperature of the most stable hairpin structure of the oligonucleotide or its complement, predicted enthalpy of the most stable hairpin structure of the oligonucleotide or its complement, predicted entropy of the most stable hairpin structure of the oligonucleotide or its complement, predicted free energy of the most stable hairpin structure of the oligonucleotide or its complement, thermodynamic partition function for intramolecular structure of the oligonucleotide or its complement.

8. A method according to Claim 1 wherein said parameter is a chemosynthetic efficiency selected from the group consisting of coupling efficiencies and overall efficiency of the synthesis of a target nucleotide sequence or an oligonucleotide probe.

9. A method according to Claim 1 wherein said parameter is a kinetic factor selected from the group consisting of steric factors calculated via molecular modeling, rate constants calculated via molecular dynamics simulations, rate constants calculated via semi-empirical kinetic modeling, associative rate constants, dissociative rate constants, enthalpies of activation, entropies of activation, and free energies of activation.

Sub B3

10. A method according to Claim 1 wherein said parameter is derived from a factor by mathematical transformation of said factor.

11. A method according to Claim 1 which comprises ranking said clustered oligonucleotides of step (d) based on the size of said clusters of oligonucleotides and selecting a subset of said clustered oligonucleotides.

12. A method according to Claim 11 wherein said subset consists of any number of oligonucleotides within said cluster of oligonucleotides.

13. A method according to Claim 11 wherein the subset of said clustered oligonucleotides are selected to statistically sample the cluster.

14. A method according to Claim 13 wherein said statistical sample consists of oligonucleotides spaced at the first quartile, median and third quartile of the cluster of oligonucleotides.

15. A method according to Claim 1 wherein said parameters are determined for said oligonucleotides by means of a computer program.

16. A method according to Claim 1 wherein said oligonucleotides are attached to a surface.

17. A method according to Claim 1 wherein said oligonucleotides are DNA.

18. A method according to Claim 1 wherein said oligonucleotides are RNA.

19. A method according to Claim 1 wherein said oligonucleotides contain chemically modified nucleotides.

20. A method according to Claim 1 wherein said target nucleotide sequence is RNA.

21. A method according to Claim 1 wherein said target nucleotide sequence is DNA.

22. A method according to Claim 1 wherein said target nucleotide sequence contains chemically modified nucleotides.

23. A method according to Claim 1 wherein said parameter is, for each oligonucleotide/target nucleotide sequence duplex, the difference between the predicted duplex melting temperature corrected for salt concentration and the temperature of hybridization of each of said oligonucleotides with said target nucleotide sequence.

24. A method according to Claim 1 wherein step (c) comprises identifying a subset of oligonucleotides within said predetermined number of unique oligonucleotides by establishing cut-off values for said parameter.

25. A method according to Claim 1 wherein said step (c) comprises identifying a subset of oligonucleotides within said predetermined number of unique oligonucleotides by converting the values of said parameter into a dimensionless number.

26. A method according to Claim 25 wherein said value is converted into a dimensionless number by determining a dimensionless score for each parameter resulting in a distribution of scores having a mean value of zero and a standard deviation of one.

27. A method according to Claim 26 which comprises optimizing a method according to calculation for said parameter based on said individual scores.

28. A method according to Claim 1 wherein step (b) comprises determining at least two parameters wherein said parameters are poorly correlated with respect to one another.

29. A method according to Claim 28 wherein said parameters are derived from a combination of factors by mathematical transformation of those factors.

30. A method according to Claim 1 wherein step (b) comprises determining two parameters at least one of said parameters being the association free energy between a subsequence within each of said oligonucleotides and its complementary sequence on said target nucleotide sequence.

31. A method according to Claim 30 wherein said subsequence is 3 to 9 nucleotides in length.

32. A method according to Claim 30 wherein said subsequence is 5 to 7 nucleotides in length.

33. A method according to Claim 30 wherein said subsequence is at least three nucleotides from the terminus of said oligonucleotides.

34. A method according to Claim 30 wherein said subsequence is at least three nucleotides from a surface to which said oligonucleotides are attached.

35. A method according to Claim 30 wherein said oligonucleotides are attached to a surface and said subsequence is at least five nucleotides from the terminus of said oligonucleotides that is attached to said surface and at least three nucleotides from the free end of said oligonucleotides.

36. A method according to Claim 30 wherein the association free energy of the members of a set of subsequences within each of said oligonucleotides is determined and said subsequence having the minimum value is identified.

37. A method according to Claim 1 which comprises including oligonucleotides that are adjacent to said oligonucleotides in said subset that are clustered along a region of said target nucleotide sequence.

38. A method according to Claim 1 which comprises (i) identifying a subset of oligonucleotides within said predetermined number of unique oligonucleotides by establishing cut-off values for each of said parameters.

39. A method according to Claim 1 which comprises determining the sizes of said clusters of step (d) by counting the number of contiguous oligonucleotides in said region of said hybridizable sequence.

40. A method according to Claim 1 which comprises determining the sizes of said clusters of step (d) by counting the number of oligonucleotides in said subset that begin in a region of predetermined length in said hybridizable sequence.

41. A method for predicting the potential of an oligonucleotide to hybridize to a complementary target nucleotide sequence, said method comprising:

(a) identifying a set of overlapping oligonucleotides from a nucleotide sequence that is complementary to said target nucleotide sequence,

(b) determining and evaluating for each of said oligonucleotides at least two parameters that are independently predictive of the ability of each of said oligonucleotides to hybridize to said target nucleotide sequence wherein said parameters are poorly correlated with respect to one another,

(c) identifying a subset of oligonucleotides within said set of oligonucleotides based on an examination of said parameters, and

(d) identifying oligonucleotides in said subset that are clustered along a region of said complementary nucleotide sequence.

42. A method according to Claim 41 which comprises ranking said oligonucleotides of step (d) based on the size of said clusters of oligonucleotides.

43. A method according to Claim 41 which comprises determining the sizes of said clusters of step (d) by counting the number of contiguous oligonucleotides in said region of said complementary sequence.

44. A method according to Claim 41 which comprises determining the sizes of said clusters of step (d) by counting the number of oligonucleotides in said subset that begin in a region of set length in said complementary sequence.

45. A method according to Claim 41 wherein said overlapping oligonucleotides are of identical length N.

46. A method according to Claim 45 wherein said overlapping oligonucleotides are spaced one nucleotide apart, said set comprising L-N+1 oligonucleotides, where L is the length of the complementary sequence.

47. A method according to Claim 41 wherein said parameters are each independently selected from the group consisting of composition factors, thermodynamic factors, chemosynthetic efficiencies and kinetic factors.

48. A method according to Claim 41 wherein said parameters are composition factors selected from the group consisting of mole fraction (G+C), percent (G+C), sequence complexity, and sequence information content.

49. A method according to Claim 41 wherein said parameters are thermodynamic factors selected from the group consisting of predicted duplex melting temperature, predicted enthalpy of duplex formation, predicted entropy of duplex formation, predicted free energy of duplex formation, predicted melting temperature of the most stable intramolecular structure of the oligonucleotide or its complement, predicted enthalpy of the most stable intramolecular structure of the oligonucleotide or its complement, predicted entropy of the most stable

intramolecular structure of the oligonucleotide or its complement, predicted free energy of the most stable intramolecular structure of the oligonucleotide or its complement, predicted melting temperature of the most stable hairpin structure of the oligonucleotide or its complement, predicted enthalpy of the most stable hairpin structure of the oligonucleotide or its complement, predicted entropy of the most stable hairpin structure of the oligonucleotide or its complement, predicted free energy of the most stable hairpin structure of the oligonucleotide or its complement, thermodynamic partition function for intramolecular structure of the oligonucleotide or its complement.

50. A method according to Claim 41 wherein any of said parameters is derived from a factor by mathematical transformation of said factor.

51. A method according to Claim 49 wherein any of said parameters is derived from a combination of factors by mathematical transformation of those factors.

52. A method according to Claim 41 wherein said parameters are chemosynthetic efficiencies selected from the group consisting of coupling efficiencies and overall efficiencies of the syntheses of a target nucleotide sequence or an oligonucleotide probe.

53. A method according to Claim 41 wherein said parameters are kinetic factors selected from the group consisting of steric factors calculated via molecular modeling, rate constants calculated via molecular dynamics simulations, rate constants calculated via semi-empirical kinetic modeling, associative rate constants, dissociative rate constants, enthalpies of activation, entropies of activation, and free energies of activation.

54. A method according to Claim 41 which comprises ranking said clustered oligonucleotides of step (d) based on the size of said clusters of oligonucleotides and selecting a subset of said clustered oligonucleotides.

55. A method according to Claim 54 wherein said subset consists of any number of oligonucleotides within said cluster of oligonucleotides.

56. A method according to Claim 54 wherein the subset of said clustered oligonucleotides are selected to statistically sample the cluster.

57. A method according to Claim 54 wherein said statistical sample consists of oligonucleotides spaced at the first quartile, median and third quartile of the cluster of oligonucleotides.

58. A method according to Claim 41 wherein said parameters are determined for said oligonucleotides by means of a computer program.

59. A method according to Claim 41 wherein said oligonucleotides are attached to a surface.

60. A method according to Claim 41 wherein said oligonucleotides are DNA.

61. A method according to Claim 41 wherein said oligonucleotides are RNA.

62. A method according to Claim 41 wherein said oligonucleotides contain chemically modified nucleotides.

63. A method according to Claim 41 wherein said target nucleotide sequence is RNA.

64. A method according to Claim 41 wherein said target nucleotide sequence is DNA.

65. A method according to Claim 41 wherein said target nucleotide sequence contains chemically modified nucleotides.

66. A method according to Claim 41 wherein said parameter is, for each oligonucleotide/target nucleotide sequence duplex, the difference between the predicted duplex melting temperature corrected for salt concentration and the temperature of hybridization of each of said oligonucleotides with said target nucleotide sequence.

67. A method according to Claim 41 wherein step (c) comprises identifying a subset of oligonucleotides within said set of oligonucleotides by establishing cut-off values for each set of parameters.

68. A method according to Claim 41 wherein said step (c) comprises identifying a subset of oligonucleotides within said set of oligonucleotides by converting the values of said parameters into a dimensionless number.

69. A method according to Claim 66 wherein said values are converted into dimensionless numbers by (a) determining a dimensionless score for each parameter resulting in a distribution of scores having a mean value of zero and a standard deviation of one and (b) calculating a combination score by evaluating a weighted average of the individual scores.

70. A method according to Claim 69 wherein step (b) comprises optimizing the weighting factors based on comparison of said individual scores to a calibration data set.

71. A method according to Claim 41 wherein step (b) comprises determining two parameters at least one of said parameters being the association free energy between a subsequence within each of said oligonucleotides and its complementary sequence on said target nucleotide sequence.

72. A method according to Claim 71 wherein said subsequence is 3 to 9 nucleotides in length.

73. A method according to Claim 71 wherein said subsequence is 5 to 7 nucleotides in length.

74. A method according to Claim 71 wherein said subsequence is at least three nucleotides from the terminus of said oligonucleotides.

75. A method according to Claim 71 wherein said oligonucleotides are attached to a surface and said subsequence is at least five nucleotides from the terminus of said oligonucleotides that is attached to said surface and at least three nucleotides from the free end of said oligonucleotides.

76. A method according to Claim 71 wherein the association free energy of the members of a set of subsequences within each of said oligonucleotides is determined and said subsequence having the minimum value is identified.

77. A method according to Claim 41 which comprises including in said evaluation oligonucleotides that are adjacent to said oligonucleotides in said subset that are clustered along a region of said target nucleotide sequence.

78. A method for predicting the potential of an oligonucleotide to hybridize to a complementary target nucleotide sequence, said method comprising:

(a) obtaining, from a nucleotide sequence complementary to said target nucleotide sequence, a set of overlapping oligonucleotides of identical length N and spaced one nucleotide apart, said set comprising L-N+1 oligonucleotides,

(b) determining and evaluating for each of said oligonucleotides the parameters: (i) the predicted melt temperature of the duplex of said oligonucleotide and said target nucleotide sequence corrected for salt concentration and (ii) predicted free energy of the most stable intramolecular structure of the oligonucleotide at the temperature of hybridization of each of said oligonucleotides with said target nucleotide sequence,

(c) identifying a subset of oligonucleotides within said set of oligonucleotides based on an examination of said parameters by establishing cut-off values for each of said parameters,

(d) ranking oligonucleotides in said subset that are clustered along a region of said complementary nucleotide sequence based on the size of said clusters of oligonucleotides, and

(e) selecting a subset of said clustered oligonucleotides.

79. A method according to Claim 78 wherein said subset consists of any number of oligonucleotides within said cluster of oligonucleotides.

80. A method according to Claim 78 wherein the subset of said clustered oligonucleotides are selected to statistically sample the cluster.

81. A method according to Claim 78 wherein said parameters are derived by mathematical transformation of the factors named in Claim 76(b).

82. A method according to Claim 78 wherein the melting temperature of step (b) is transformed by subtracting the temperature of hybridization.

83. A method according to Claim 78 which comprises determining the sizes of said clusters of step (d) by counting the number of contiguous oligonucleotides in said region of said complementary sequence.

84. A method according to Claim 78 wherein said statistical sample consists of oligonucleotides spaced at the first quartile, median and third quartile of the cluster of oligonucleotides.

85. A method according to Claim 78 wherein said parameters are determined for said oligonucleotides by means of a computer program.

86. A method according to Claim 78 wherein said oligonucleotides are attached to a surface.

87. A method according to Claim 78 wherein said oligonucleotides are DNA.

88. A method according to Claim 78 wherein said oligonucleotides are RNA.

89. A method according to Claim 78 wherein said oligonucleotides contain chemically modified nucleotides.

90. A method according to Claim 78 wherein said target nucleotide sequence is RNA.

91. A method according to Claim 78 wherein said target nucleotide sequence is DNA.

92. A method according to Claim 78 wherein said target nucleotide sequence contains chemically modified nucleotides.

93. A method according to Claim 68 wherein the following equations are used for converting the values of said parameters into a dimensionless number:

$$s_{i,x} = \frac{x_i - \langle x \rangle}{\sigma_{\{x\}}},$$

where $s_{i,x}$ is the dimensionless score derived from parameter x calculated for oligonucleotide i, x_i is the value of parameter x calculated for oligonucleotide i, <x> is the average of parameter x calculated for all of the oligonucleotides under consideration for a given nucleotide sequence target, and $\sigma_{\{x\}}$ is the standard deviation of parameter x calculated for all of the oligonucleotides under consideration for a given nucleotide sequence target, and is given by the equation

$$\sigma_{\{x\}} = \sqrt{\frac{\sum_{j=1}^{L-N+1}\left(x_j - \langle x\rangle\right)^2}{L-N}}.,$$

where the target sequence is of length *L* and the oligonucleotides are of length *N*.

94. A method according to Claim 68 wherein a combination score S_i is calculated by evaluating a weighted average of the individual values of the dimensionless scores $s_{i,x}$ by the equation:

$$S_i = \sum_{\{x\}} q_x \, s_{i,x} \,,$$

where q_x is the weight assigned to the score derived from parameter *x*, the individual values of q_x are always greater than zero, and the sum of the weights q_x is unity.

95. A method according to Claim 78 where clustering is determined by calculating a moving window-averaged combination score $\langle S_i \rangle$ for the *i*th probe by the equation:

$$\langle S_i \rangle = \frac{1}{w} \sum_{j=i-\frac{w-1}{2}}^{i+\frac{w-1}{2}} S_j \,, \; w = \text{an odd integer}.,$$

where *w* is the length of the window for averaging, and then applying a cutoff filter to the value of $\langle S_i \rangle$.

96. A method according to Claim 94 wherein optimization of the weights q_x is performed by varying the values of the weights so that the correlation coefficient $\rho_{\{\langle Si\rangle\},\{Vi\}}$ between the set of window-averaged combination scores

$\{<S_i>\}$ and a set of calibration experimental measurements $\{V_i\}$ is maximized. The correlation coefficient $\rho_{\{<S_i>\},\{V_i\}}$ is calculated from the equation

$$\rho_{x,y} = \frac{\text{Covariance}(x,y)}{\sqrt{\text{Variance}(x)\text{Variance}(y)}},$$

where $x=<S_i>$, $y=V_i$ and the Covariance (x,y) is defined by

$$\text{Covariance}(x,y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y).$$

The quantities μ_x and μ_y are the averages of the quantities x and y, while the variances are the squares of the standard deviations.

97. A method according to Claim 95 wherein the cutoff filter selects the lowest values of the window-averaged combination score $<S_i>$ and the clustered probes so identified are predicted to exhibit low hybridization efficiency.

98. A computer based method for predicting the potential of an oligonucleotide to hybridize to a target nucleotide sequence, said method comprising:

(a) identifying under computer control a predetermined number of unique oligonucleotides within a nucleotide sequence that is hybridizable with said target nucleotide sequence, said oligonucleotides being chosen to sample the entire length of said nucleotide sequence,

(b) under computer control, determining and evaluating for each of said oligonucleotides a value for at least one parameter that is independently predictive of the ability of each of said oligonucleotides to hybridize to said target nucleotide sequence and storing said parameter values,

(c) identifying under computer control, from said stored parameter values, a subset of oligonucleotides within said predetermined number of unique oligonucleotides based on an examination of said parameter, and

(d) identifying under computer control oligonucleotides in said subset that are clustered along a region of said nucleotide sequence that is hybridizable to said target nucleotide sequence.

99. A method according to claim 98 wherein the identified subset of oligonucleotide sequences is electronically transferred to an oligonucleotide array manufacturing system.

100. A computer system for conducting a method for predicting the potential of an oligonucleotide to hybridize to a target nucleotide sequence, said system comprising:

(a) input means for introducing a target nucleotide sequence into said computer system,

(b) means for determining a number of unique oligonucleotide sequences that are within a nucleotide sequence that is hybridizable with said target nucleotide sequence, said oligonucleotide sequences being chosen to sample the entire length of said nucleotide sequence,

(c) memory means for storing said oligonucleotide sequences,

(d) means for controlling said computer system to carry out a determination and evaluation for each of said oligonucleotide sequences a value for at least one parameter that is independently predictive of the ability of each of said oligonucleotide sequences to hybridize to said target nucleotide sequence,

(e) means for storing said parameter values,

(f) means for controlling said computer to carry out an identification from said stored parameter values a subset of oligonucleotide sequences within said number of unique oligonucleotide sequences based on an examination of said parameter,

(g) means for storing said subset of oligonucleotides,

(h) means for controlling said computer to carry out an identification of oligonucleotide sequences in said subset that are clustered along a region of said nucleotide sequence that is hybridizable to said target nucleotide sequence.

(i) means for storing said oligonucleotide sequences in said subset, and

(j) means for outputting data relating to said oligonucleotide sequences in said subset.

101. A computer system according to claim 100 wherein the identified subset of oligonucleotide sequences is electronically transferred to an oligonucleotide array manufacturing system.